DUE: December 17th, 2022

Analyzing Factors Contributing to Mental Health Outcomes in Children

Exploratory Data Analysis

Team #7 | Alex Martinez, Ikram Salim, Lucero Huayhua, Marylyn Rogel, Ugochi Ejiogu, Zach Jackson

Team Member	Contribution
Alex M.	‘Single Variable Distributions’, ‘Null/Missing Data’, team python notebook
Ikram S.	Team python notebook
Lucero H.	‘Correlations’, analysis of Health Care Services- and Health Insurance-related variables, team python notebook
Marylyn R.	Team python notebook
Ugochi E.	‘Columns of Interest’, ‘Null/Missing Data’, ‘Ordinal Data’, analysis of ACEs-related variables
Zach J.	‘Categorical Data’, Created Team 7 Google account for shared code/notebook sharing

Contents

Columns of Interest

For this exploratory data analysis report, we will be working with a selection of 49 columns from the US Census 2021 National Survey of Children's Health topical dataset. The insight gained from exploring our data will guide us in conducting our extended analyses in the future.

Our capstone project aims to evaluate what factors most contribute to mental health outcomes in children. In order to answer this question, we are exploring our data with basic statistics and visualization of each variable and searching for possible relationships between variables to better understand our data. Our columns of interest include the following:

Variable	Type	Description
BIRTH_YR_F	FLOAT	BIRTH_YR Data Quality Flag
FIPSST	OBJECT	State FIPS Code
FORMTYPE	OBJECT	Form Type
HEADACHE	FLOAT	Headaches
K2Q01	FLOAT	General Health
K2Q30A	FLOAT	Learning Disability
K2Q31A	FLOAT	ADD/ADHD
K2Q32A	FLOAT	Depression
K2Q33A	FLOAT	Anxiety
K2Q34A	FLOAT	Behavior Problems
K2Q37A	FLOAT	Speech Disorder
C4Q04	FLOAT	Frustrated In Efforts to Get Service
K4Q22_R	FLOAT	Mental Health Professional Treatment
K4Q23	FLOAT	Emotions Concentration Behavior Medication
K4Q27	FLOAT	Needed Health Care Not Received
OVERWEIGHT	FLOAT	Doctor Identified as Overweight
CURRCOV	FLOAT	Health Insurance Coverage - Currently Covered (Use CURRINS)
INSGAP	FLOAT	Health Insurance Coverage - Past 12 Months
AVOIDCHG	FLOAT	Past 12 Months - Avoided Changing Jobs to Maintain Health Insurance
K3Q25	FLOAT	Problems Paying for Medical or Health Care
BORNUSA	FLOAT	Born in the United States
K8Q35	FLOAT	Someone to Turn To for Emotional Support
ACE1	FLOAT	Hard to Cover Basics Like Food or Housing
ACE10	FLOAT	Child Experienced - Treated Unfairly Because of Race

ACE3	FLOAT	Child Experienced - Parent or Guardian Divorced
ACE4	FLOAT	Child Experienced - Parent or Guardian Died
ACE5	FLOAT	Child Experienced - Parent or Guardian Time in Jail
ACE6	FLOAT	Child Experienced - Adults Slap, Hit, Kick, Punch Others
ACE7	FLOAT	Child Experienced - Victim of Violence
ACE8	FLOAT	Child Experienced - Lived with Mentally Ill
ACE9	FLOAT	Child Experienced - Lived with Person with Alcohol/Drug Problem
FOODSIT	FLOAT	Food Situation In Household - Past 12 Months
GOFORHELP	FLOAT	Know Where to Go For Help
K10Q12	FLOAT	Neighborhood - Park or Playground
K10Q13	FLOAT	Neighborhood - Recreation Center
K10Q40_R	FLOAT	Child is Safe In Neighborhood
A1_AGE	FLOAT	Adult 1 - Age in Years
A1_BORN	FLOAT	Adult 1 - Where Born
A1_EMPLOYED	FLOAT	Adult 1 - Current Employment Status
A1_MARITAL	FLOAT	Adult 1 - Marital Status
A1_MENTHEALTH	FLOAT	Adult 1 - Mental or Emotional Health
A2_AGE	FLOAT	Adult 2 - Age in Years
A2_BORN	FLOAT	Adult 2 - Where Born
A2_EMPLOYED	FLOAT	Adult 2 - Current Employment Status
A2_MARITAL	FLOAT	Adult 2 - Marital Status
A2_MENTHEALTH	FLOAT	Adult 2 - Mental or Emotional Health
A2_RELATION	FLOAT	Adult 2 - How Related to Child
HIGRADE	FLOAT	Highest Level of Education among Reported Adults
HOUSE_GEN	FLOAT	Parental Nativity

Preliminary Insights

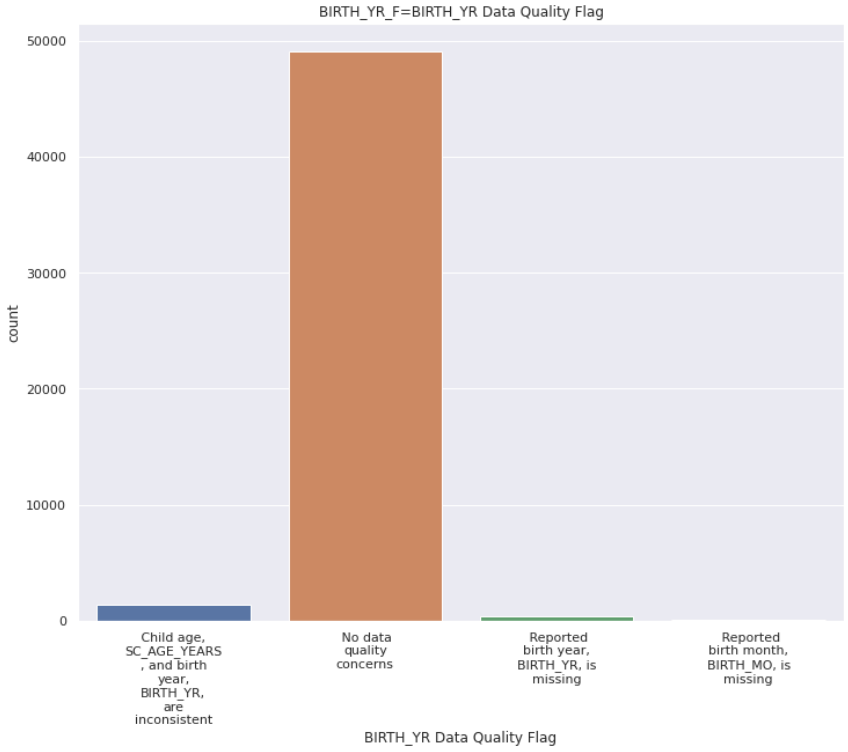
Null/Missing Data

A majority of the columns have minimal null values (< 5%). The variable ‘K3Q25 - Problems Paying for Medical or Health Care’ is missing almost 30% of its data and most of the columns related to Adult 2 (‘A2’) are missing ~18% of its data. Around 14.2% of A2 relationships are coded as “There is only one primary adult caregiver in the household for this child”. This means that these single-parent households can be expected to have about 14.2% of A2 variables missing. The BIRTH_YR_F is a data quality flag that cross-references various data columns for

consistency. A bar graph of this is included below. A vast majority of the data has no data quality issues.

Variable	Percent Null Values
BIRTH_YR_F	0.00%
FIPSST	0.00%
FORMTYPE	0.00%
HEADACHE	0.22%
K2Q01	0.22%
K2Q30A	0.30%
K2Q31A	0.56%
K2Q32A	0.33%
K2Q33A	0.39%
K2Q34A	0.35%
K2Q37A	0.38%
C4Q04	0.57%
K4Q22_R	0.50%
K4Q23	2.60%
K4Q27	0.40%
OVERWEIGHT	0.46%
CURRCOV	0.43%
INSGAP	0.93%
AVOIDCHG	1.18%
K3Q25	29.64%
BORNUSA	0.98%
K8Q35	2.29%
ACE1	2.19%
ACE10	4.10%
ACE3	3.71%
ACE4	3.89%
ACE5	3.98%
ACE6	4.03%
ACE7	4.06%
ACE8	4.07%
ACE9	3.99%

FOODSIT	2.78%
GOFORHELP	3.31%
K10Q12	2.86%
K10Q13	2.97%
K10Q40_R	3.15%
A1_AGE	3.16%
A1_BORN	3.06%
A1_EMPLOYED	3.73%
A1_MARITAL	3.60%
A1_MENTHEALTH	3.52%
A2_AGE	17.96%
A2_BORN	18.78%
A2_EMPLOYED	18.16%
A2_MARITAL	18.31%
A2_MENTHEALTH	18.15%
A2_RELATION	3.48%
HIGRADE	0.00%
HOUSE_GEN	1.96%



Numerical Data

The dataset includes the following numerical variables for analysis:

Variable	Description	Response Code
A1_AGE	Adult 1 - Age in Years	18-75 or older
A2_AGE	Adult 2 - Age in Years	18-75 or older

(see visualizations under Single Variable Distributions)

Ordinal Data

The dataset includes the following numerical variables that are treated ordinally for analysis:

Variable	Description	Response Code
K2Q01	General Health	1 = Excellent | 2 = Very Good | 3 = Good | 4 = Fair | 5 = Poor
C4Q04	Frustrated in Efforts to Get Service	1 = Never | 2 = Sometimes | 3 = Usually | 4 = Always
ACE1	Hard to Cover Basics Like Food or Housing	1 = Never | 2 = Rarely | 3 = Somewhat Often | 4 = Very Often
GOFORHELP	Know Where to Go for Help	1 = Definitely Agree | 2 = Somewhat Agree | 3 = Somewhat Disagree | 4 = Definitely Disagree
K10Q40_R	Child is Safe in Neighborhood	1 = Definitely Agree | 2 = Somewhat Agree | 3 = Somewhat Disagree | 4 = Definitely Disagree
A1_MENTHEALTH	Adult 1 - Mental or Emotional Health	1 = Excellent | 2 = Very Good | 3 = Good | 4 = Fair | 5 = Poor
A2_MENTHEALTH	Adult 2 - Mental or Emotional Health	1 = Excellent | 2 = Very Good | 3 = Good | 4 = Fair | 5 = Poor

(see notebook for visualizations)

Categorical Data

The dataset includes the following categorical variables for analysis:

Variable	Description	Response Code
FIPSST	State FIPS Code	1-56 (link to response code)
FORMTYPE	Form Type (Operational)	T1: children aged 0 through 5

		T2: children aged 6 through 11 T3: children aged 12 through 17
BIRTH_YR_F	BIRTH_YR Data Quality Flag	0 = No data quality concerns 1 = Child age, SC_AGE_YEARS, and birth year, BIRTH_YR, are inconsistent 2 = Reported birth year, BIRTH_YR, is missing 3 = Reported birth month, BIRTH_MO, is missing
K4Q22_R	Mental Health Professional Treatment in past 12 months	1 = Yes 2 = No, but this child needed to see a mental health professional 3 = No, this child did not need to see a mental health professional
INSGAP	Health Insurance Coverage - Past 12 Months	1 = Insured all 12 months 2 = Insured during the past 12 months but with gaps in coverage 3 = No coverage past 12 months
FOODSIT	Food Situation in Household - Past 12 Months	1 = We coud always afford to eat good nutritious meals 2 = We could always afford enough to eat but not always the kinds of food we should eat 3 = Sometimes we could not afford enough to eat 4 = Often we could not afford enough to eat
A1_BORN	Adult 1 - Where Born	1 = In the United States 2 = Outside of the United States
A1_EMPLOYED	Adult 1 - Current Employment Status	1 = Employed full-time 2 = Employed part-time 3 = Working WITHOUT pay 4 = Not employed but looking for work 5 = Not employed but not looking for work
A1_MARITAL	Adult 1 - Marital Status	1 = Married 2 = Not married, but living with a partner 3 = Never Married 4 = Divorced 5 = Separated 6 = Widowed
A2_BORN	Adult 2 - Where Born	1 = In the United States 2 = Outside of the United States
A2_MARTIAL	Adult 2 - Marital Status	1 = Married 2 = Not married, but living with a partner 3 = Never Married

		4 = Divorced 5 = Separated 6 = Widowed
A2_RELATION	Adult 2 -How Related to Child	1 = Biological or Adoptive Parent 2 = Step-parent 3 = Grandparent 4 = Foster Parent 6 = Other: Relative 7 = Other: Non-Relative
HIGRADE	Highest Level of Education among Reported Adults	1 = Less than high school 2 = High school (including vocational, trade, or business school) 3 = More than high school
HOUSE_GEN	Parental Nativity	1 = 1st generation household [Child is born outside the United States and all reported parents are born outside the United States. At least one parent must be reported as born outside the United States.] 2 = 2nd generation household [Child is born in the United States and at least one parent is born outside the United States OR child is born outside the United States, one parent is born in the United States and one parent is born outside the United States.] 3 = 3rd+ generation [All parents in the household are born in the United States] 4 = Other [Child is born in the United States, parents are not listed.]

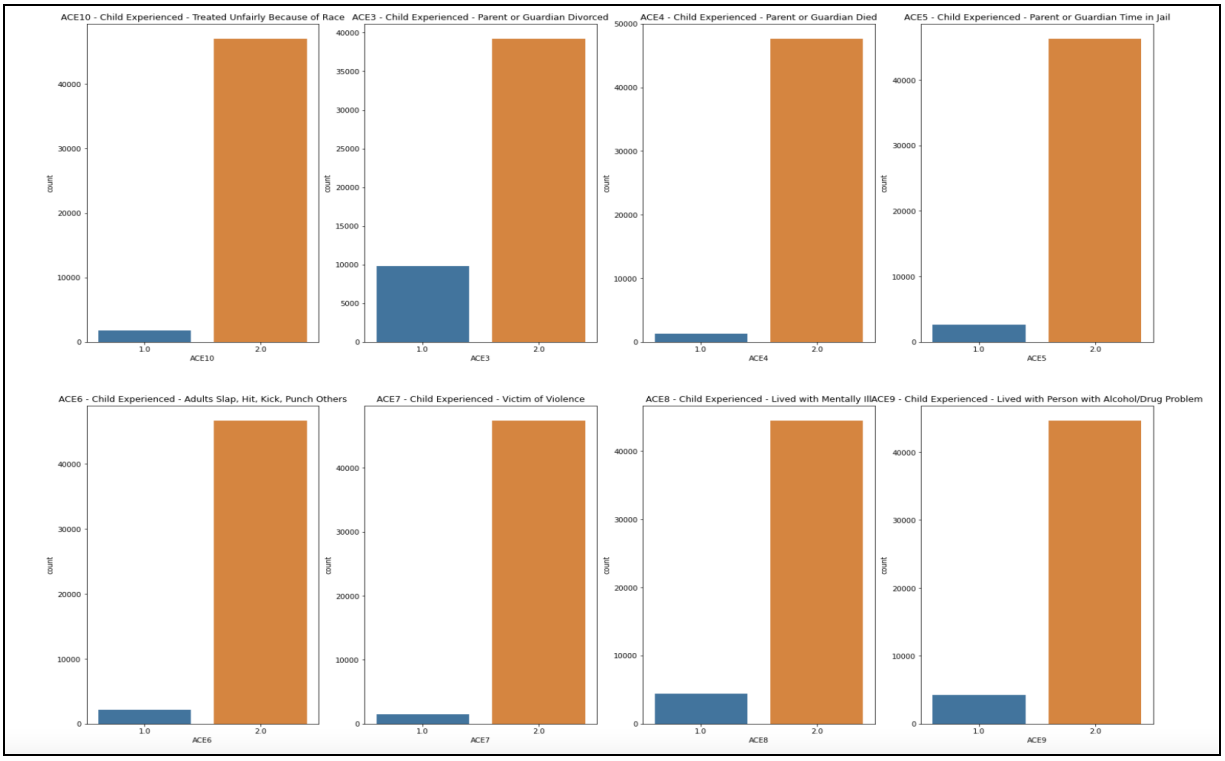
(see notebook for visualizations)

There were more categorical variables that upon further discussion should be added (i.e., A1_RELATION and A2_EMPLOYED)

Nominal Data

The rest of the variables for analysis have yes (1) / no (2) responses:

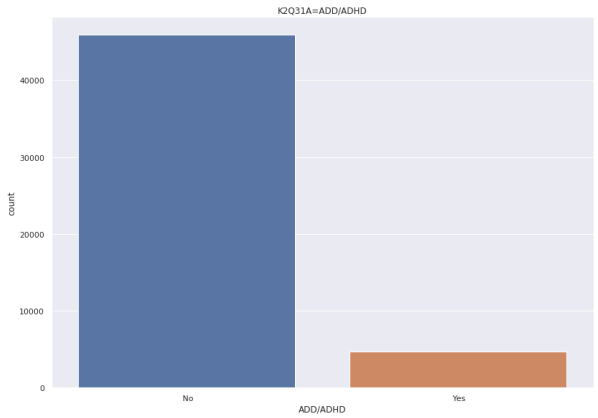
For example, the value counts from the columns related to Adverse Childhood Experiences (ACEs) are plotted below.

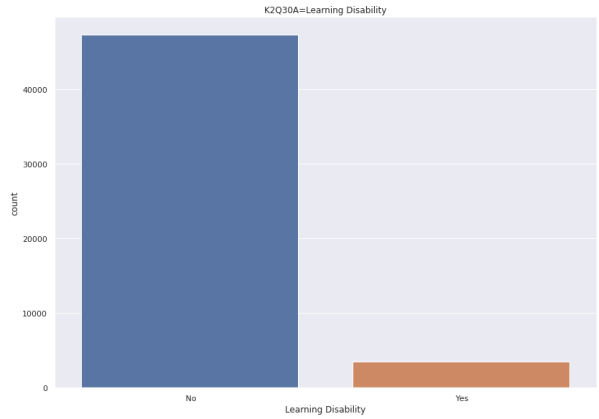


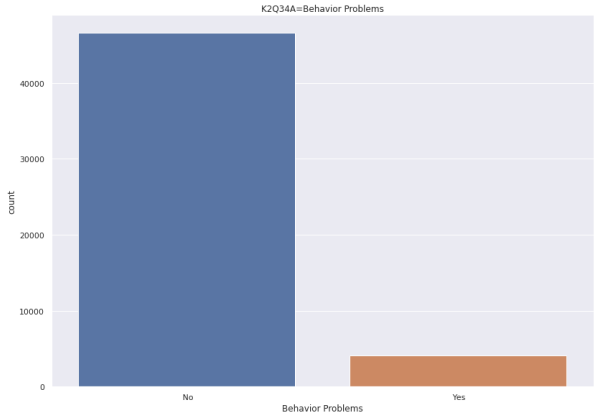
Single Variable Distributions

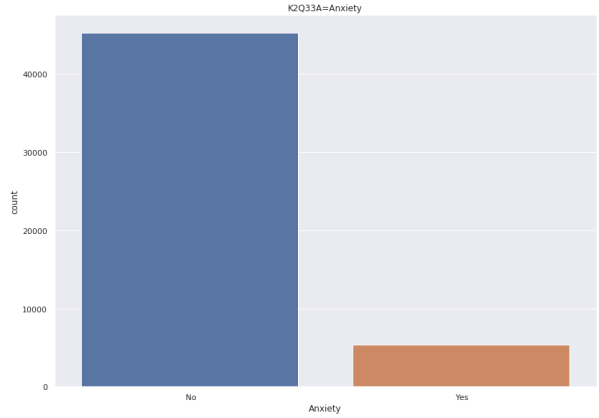
We plotted bar graphs for each of the variables of interest. The data is fairly one-sided for the nominal values and ordinal data. As the plot above shows, the orange bars are much larger. Moreover, for several of the child health variables, the factors that may necessitate additional mental health help are in the minority. This is the case for a majority of the variables. On the other hand, some variables, especially those non-dichotomous questions have a more diverse distribution. For example, the Adult 1 mental or emotional health variable includes a scale from poor to excellent. The data is not totally dominated by the “Excellent” column. The distribution of the Adult ages seems to be approximately normal with a slightly fatter right tail. Missing ages were coded as -1. At the geographic level, most of the states were represented approximately equally. You may notice in the final graph that 41, 55, 08, and 39 stand out. These correspond to Oregon, Wisconsin, Colorado, and Ohio having significantly more data included in this data.

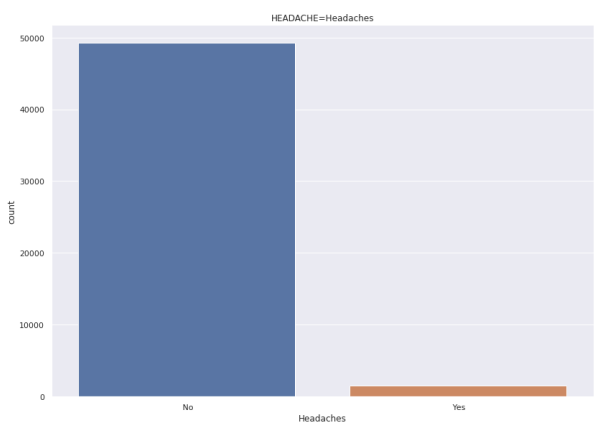
Additional single-variable visualizations are available in our team notebook.

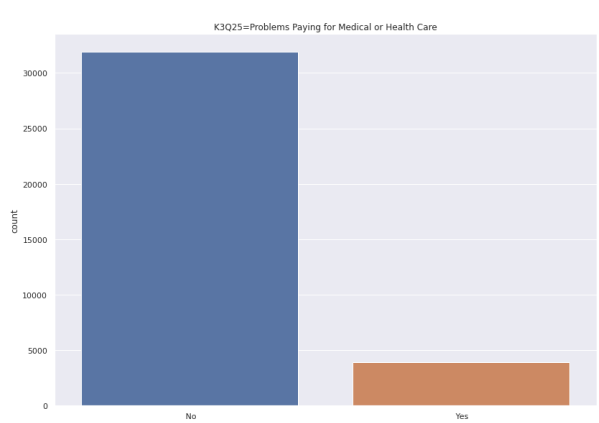

K2Q31A=ADD/ADHD
count
40000
30000
20000
10000
0
No
Yes
ADD/ADHD

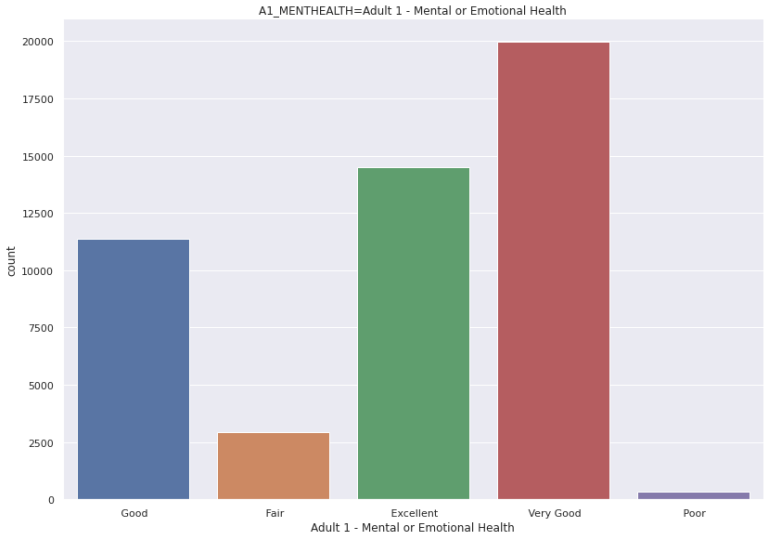

K2Q30A=Learning Disability
count
40000
30000
20000
10000
0
No
Yes
Learning Disability

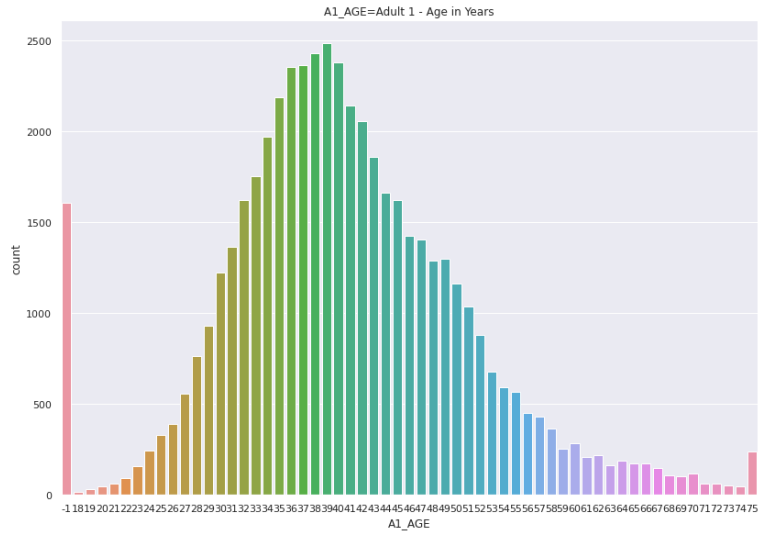

K2Q34A=Behavior Problems
count
40000
30000
20000
10000
0
No
Yes
Behavior Problems

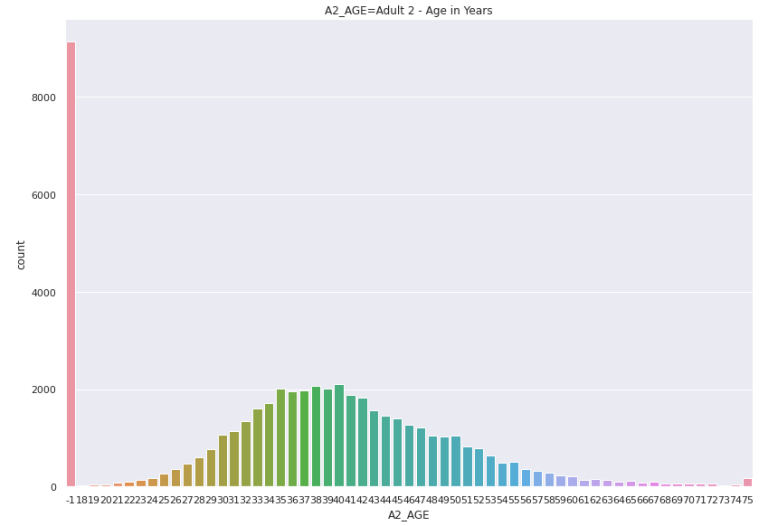

K2Q33A=Anxiety
count
40000
30000
20000
10000
0
No
Yes
Anxiety

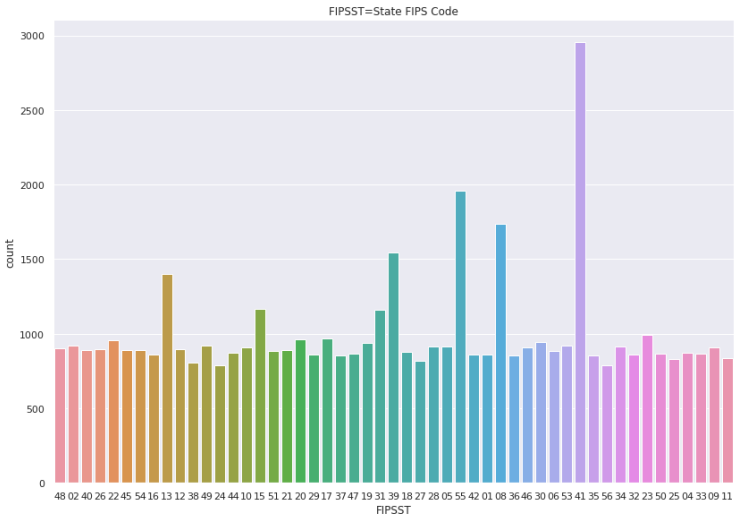

HEADACHE=Headaches
count
50000
40000
30000
20000
10000
0
No
Yes
Headaches


K3Q25=Problems Paying for Medical or Health Care
count
30000
25000
20000
15000
10000
5000
0
No
Yes
Problems Paying for Medical or Health Care


A1_MENTHEALTH=Adult 1 - Mental or Emotional Health
count
0
2500
5000
7500
10000
12500
15000
17500
20000
Good
Fair
Excellent
Very Good
Poor
Adult 1 - Mental or Emotional Health


A1_AGE=Adult 1 - Age in Years
count
0
500
1000
1500
2000
2500
A1_AGE

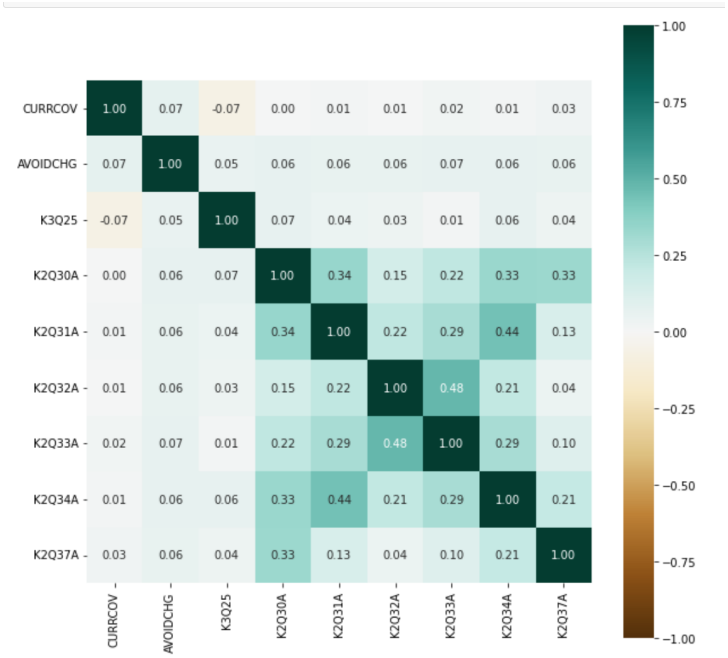

A2_AGE=Adult 2 - Age in Years
count
0
2000
4000
6000
8000
A2_AGE



Correlations

In this correlation analysis we included Healthcare Services related variables (C4Q04, K4Q22_R, K4Q23, K4Q27, OVERWEIGHT) and depression, anxiety, learning disability, ADHD, behavior problems, and speech disorder variables.

In this correlation analysis, we compared Health Insurance related variables (CURRCOV, AVOIDCHG, K3Q25) against depression, anxiety, learning disability, ADHD, behavior problems, and speech disorder variables. There are no strong correlations between Health Insurance related variables and the mentioned conditions.



In this correlation analysis, we compared the ACEs-related variables with depression, anxiety, learning disability, ADHD, behavior problems, and speech disorder variables. There are no strong correlations between ACEs and the mentioned conditions.